



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/468,785	12/21/1999	TIMOTHY L. HUMPHREY	P64616US0	4093

7590 05/07/2004

JACOBSON, PRICE, HOLMAN & STERN, PLLC
400 Seventh Street, N. W.
Washington, DC 20004

EXAMINER

SPOONER, LAMONT M

ART UNIT

PAPER NUMBER

2654

DATE MAILED: 05/07/2004

9

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	09/468,785	HUMPHREY ET AL.
Examiner	Art Unit	
Lamont M Spooner	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) Responsive to communication(s) filed on 09 February 2004.
 2a) This action is **FINAL**. 2b) This action is non-final.
 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) Claim(s) 1-4, 16-19, 31-34 and 46-60 is/are pending in the application.
 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
 5) Claim(s) _____ is/are allowed.
 6) Claim(s) 1-4, 16-19 and 31-34 is/are rejected.
 7) Claim(s) 48-50, 53-55 and 58-60 is/are objected to.
 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) The specification is objected to by the Examiner.
 10) The drawing(s) filed on 21 December 1999 is/are: a) accepted or b) objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) Notice of References Cited (PTO-892)
 2) Notice of Draftsperson's Patent Drawing Review (PTO-948)
 3) Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
 Paper No(s)/Mail Date _____.
 4) Interview Summary (PTO-413)
 Paper No(s)/Mail Date _____.
 5) Notice of Informal Patent Application (PTO-152)
 6) Other: _____.

DETAILED ACTION

Information Disclosure Statement

1. The listing of references in the specification is not a proper information disclosure statement. 37 CFR 1.98(b) requires a list of all patents, publications, or other information submitted for consideration by the Office, and MPEP § 609 A(1) states, "the list may not be incorporated into the specification but must be submitted in a separate paper." Therefore, unless the references have been cited by the examiner on form PTO-892, they have not been considered.
2. The information disclosure statement filed 6/11/2000 fails to comply with 37 CFR 1.98(a)(1), which requires a list of all patents, publications, or other information submitted for consideration by the Office. It has been placed in the application file, but the information referred to therein has not been considered.

Response to Arguments

3. Applicant's arguments filed 02/09/2004 have been fully considered but they are not persuasive.

- In response to applicant's argument on p. 16 lines 5-9 that the texts being compared come from two different documents and based on the comparison process, a content score is calculated for each text unit and at least one text unit is selected from the citing documents' contexts as constituting the RFC on the basis of scoring.

Lawrence et al. teaches comparing text between different documents to determine relatedness (C.16.lines 32-37, C.17.lines 42-44-the word vector

calculation compares the text between different documents, which is the analysis of the content of different documents, the Term Frequency time Inverse Document Frequency (TDIDF) determines the content score for each text unit based on a comparison process). The context, which is determined as the RFC, is displayed based on the analyzed contents of the contexts, (C.9.lines 29-34-the context, which is interpreted to include any text before or after and the citing instance, is extracted based on the analyzed content of the contexts of the citing document, the analysis including frequency calculation, stop word removal, word stemming and word vector analysis, C.16.lines 35-37, C.16.lines 61, 62-“the relatedness” is identified by the context as the RFC).

- In response to applicant’s arguments, p. 16 lines 15-28 that the step of analyzing the content of the context is not taught by Lawrence et al., including comparing portions of text taken from two documents to identify content words that are common to both documents.

Lawrence et al. teaches using word vector comparison (C.16.lines 35-37), which compares portions of text taken from two documents, which is a part of the analysis of the content of the contexts, and Term Frequency time Inverse Document Frequency (TFIDF) (C.19.line 11) as a scoring method gives weight to each content word in the document, as well as accounting for the same word occurring in different documents (C.16.lines 16-18, 32-35), which are included in the database as common content words (C.8.lines 13-17).

- In response to applicant's argument, p. 16 line 21 and p. 17 lines 1-3, that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., "the specific reason for which the article was citing through comparison with one or more citing instances appearing in the other citing documents.") are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).

Lawrence et al. teaches determining the relatedness of documents (C.16.lines 60-67), which is interpreted as the reason for which the article was cited, through comparison with citing instances on other citing documents (C.17.lines 1-4).

- In response to applicant's arguments, p.17 lines 8-12, that Fig 6 of Lawrence et al. does not suggest an automated comparison of the words within the two text sets to identify commonality, or the automated scoring of text units on the basis of common content words.

Lawrence et al. teaches using word vector comparison (C.16.lines 35-37) between different documents and TFIDF on every content word (C.16. lines 31-34, C.19.line 11), which includes a scoring for words that are common to different documents, in an automated process (C.1.lines 12-17). Therefore the context presented by Lawrence et al. in Fig. 6 is interpreted as the reasons for citing.

- In response to applicant's question of the comparison of two texts in generation of a content word list, p. 17 lines 17-19.

The TFIDF (C.19.line 11), which includes a comparison of common words in the context of the citing document and the cited document, and word frequency are recorded in the new database (C.8.lines 12-16) is interpreted as the content word list, wherein the content of the context has been analyzed, then selected and displayed (C.8.lines 12-16, C.16.lines 35-37).
- In response to applicant's arguments, p. 18 lines 2-6, that Lawrence et al. does not teach requiring that words entered into the new database be included in the contexts of at least two citing documents or in the context of one citing document and the cited document, nor requiring or suggesting such document cross-comparison as a predicate for inclusion in a subsequent listing.

Lawrence et al. teaches the content words that are common to different documents, determined by word frequency information (TFIDF-which takes into account the common words between different documents) (C.16.lines 31-34, C.19.line 11), including at least two citing documents (C.16.lines 62-65), and other content words are included in the "new database" as the content list (C.8.lines 12-16), the content analysis is performed (as a predicate)/before the content word can be included in the new database (C.8.lines 12-16).
- In response to applicant's arguments against the references individually, p.18 lines 9-18, one cannot show nonobviousness by attacking references

individually where the rejections are based on combinations of references. See *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981); *In re Merck & Co.*, 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986).

- In response to applicant's argument that there is no suggestion to combine the references, p. 18 lines 19-21, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988) and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, Pedersen et al. provides a scoring method for individual paragraphs and sentences (C.3.lines 16-18), which enhances and improves the mode of operation in parsing the documents, for Lawrence et al.

4. Applicant's arguments with respect to claims 46-60, p. 19 lines 11-21, p. 20 lines 1-10, have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 102

5. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States

only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

6. Claims 1, 16 and 31 are rejected under 35 U.S.C. 102(e) as being anticipated by Lawrence et al. (US Patent No. 6,289,342 filed 5/20/98).

As per **claims 1, 16 and 31**, Lawrence et al. discloses:

an automated method of designating text, taken from a set of citing documents, as reasons for citing (RFC) a cited document that are associated with respective citing instances of a cited document, the method comprising:

obtaining contexts of the citing instances in the respective citing documents, each context (C.9.lines 30-35-“This allows extraction of the context of the citations”) including a text unit that includes the citing instance and a text unit that is near the citing instance (Fig 6 “very powerful [8], especially...”, C.19.lines 11, 12- The Term Frequency Inverse Document Frequency (TFIDF) and Identical Citation Grouping (ICG) extraction includes the citing instance and a text unit near the citing instance);

analyzing the content of the contexts, said step of analyzing including calculating a content score for each text unit based on text unit content words that are common to at least two of the citing documents’ contexts (C.16. lines 16-18, 35-37- which compares portions of text taken from two documents, which is a part of the analysis of the content of the contexts, and Term Frequency times Inverse Document Frequency (TFIDF), C.19.line 11, as a scoring method gives weight to each content word in the document, as well as accounting for the same word occurring in different documents , which are included in the new database as common content words, C.8.lines 12-16) or to at least one citing document’s context and said cited document; and

selecting from the citing instances' context, at least one text unit that constitutes the RFC, based on the analyzed content of the contexts (Fig 6-includes at least one text unit that constitutes the RFC, C.9.lines 30-35, C.16.lines 35-37- which compares portions of text taken from two documents, which is a part of the analysis of the content of the contexts).

Claim Rejections - 35 USC § 103

7. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

8. Claims 2-4, 17-19 and 32-34 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lawrence et al. in view of Pedersen et al. (US Patent No 5,638,543 filed June 3, 1993).

Lawrence et al. and Pedersen et al. are analogous art because they are both automated document manipulation methodology related.

As per **claims 2, 17, and 32** Lawrence discloses:

an automated method of designating text, taken from a set of citing documents, as reasons for citing (RFC) a cited document, said RFC being associated with respective citing instances of a citing document, the method comprising:

inputting text from the citing documents (C.9.lines 6, 7);

obtaining contexts of the citing instances in the respective citing documents, each context including: a sentence that includes the citing instance and at least one sentence that is near the citing instance (C.11.lines 33-37, Fig 6 "1. INTRODUCTION Second-order recurrent networks have proven to be very powerful [8], especially when trained using complete back propagation through time [1, 6, 14]. It has also been demonstrated by ...");

generating a content word list containing content words that are in at least two of the citing documents' contexts (C.16.lines 16, 17, 31-34, C.19.line 11, the TFIDF, which includes a comparison of common words in the context of the citing document and the cited document, C.16.lines 5-10-document of interest is interpreted as a cited document, C.16.lines 61, 62-the related documents are interpreted to include the citing document and cited document, and word frequency are recorded in the new database, C.8.lines 12-16, containing the content word list, wherein the content of the context has been analyzed, then selected and displayed C.8.lines 12-16) or that are in at least one citing document's context and said cited document.

calculating, for the sentences in the citing documents' contexts, respective content scores that are based on the frequency counts of the content words that are recited in the respective sentences (C.9.lines 9-12-determines the frequency count, C.16.lines 31-34, C.19.line 11-the TFIDF which incorporates an account of common words , C.8.lines 12-16-the new database which holds the content list); and

selecting, from the citing documents' contexts, at least one sentence that constitute the RFC (Fig 6 "It has also been demonstrated...").

Lawrence does not disclose:

dividing the citing documents' texts to define paragraphs, and dividing the paragraphs to define sentences;
selecting the sentences that constitute the RFC based on the calculated content scores.

However, as it is well known in the art, Pedersen et al. teaches:

dividing the documents' text texts to define paragraphs, and dividing the paragraphs to define sentences (C.3.lines 4-9).
selecting the sentences based on the calculated content scores (C.3.lines 16-18).

Therefore it would have been obvious to a person skilled in the art to combine Lawrence et al. with Pedersen et al. The motivation for doing so would have been to select text based on scoring from the citing documents.

As per **claims 3, 18 and 33** Lawrence et al. discloses all of the limitations of claim 1 upon which claim 3 depends. Lawrence et al. further discloses:

the step of analyzing the content includes;
generating a content word list based on the content words that are included in the contexts of at least two of the citing documents (C.8.lines 12-16-the new database contains the content words common to the contexts of at least two of the citing documents, C.17.lines 1-4-which contains the two citing documents, as well as additional content words), and assigning each of said content words a frequency count which is used in calculating the content score (C.16.lines 31-34, C.19.line 11-TFIDF-the

TF is a frequency count assigned to each content word which is used in calculating the weight, which is interpreted as part of calculating a content score).

As per **claims 4, 19 and 34**, Lawrence et al. discloses all of the limitations of claim 1 upon which claim 4 depends. Lawrence et al. further discloses:

the method, wherein the step of analyzing the content includes; generating a content word list based on the content words that are included both in the cited document itself (C.8.lines 52-54-the cited document) and in the context of at least one citing document, (C.19.line 11, the TFIDF, which includes a comparison of common words in the context of the citing document and the cited document, C.16.lines 5-10-document of interest is interpreted as a cited document, C.16.lines 61, 62-the related documents are interpreted to include the citing document and cited document, and word frequency are recorded in the new database, C.8.lines 12-16, containing the content word list), and assigning each of said content words a frequency count which is used in calculating the content score (C.16.lines 31-34, C.19.line 11-In the TFIDF, TF is a frequency count assigned to each content word which is used in calculating the weight, which is interpreted as part of calculating a content score).

As per **claims 46, 51 and 56**, Lawrence et al. and Pedersen et al. disclose all of the limitations of claim 2, upon which claim 46 depends. Lawrence et al. further discloses:

associating paragraphs from the documents (C.4.lines 13-19-the bodies of text/paragraphs between documents are associated with one another in determining similarity).

processing text in the associated paragraphs to eliminate noise words that convey little information about paragraph content (C.4.lines 51-55).

determining common words that are not eliminated by the processing step and that are found in at least two paragraphs (C.4.lines 42, 43-the common words are in different documents, which places them in at least two paragraphs).

forming the content word list to include the common words linked to respective frequency counts (C.8.lines 12-16, C.19.line 11-the content word list included in the new database contains the TF of all content words including common words and the linked frequencies therewith).

Lawrence et al. does not disclose:

tallying frequency counts that indicate respective numbers of paragraphs within which the common words are encountered, said frequency counts indicating a degree of relevance for respective common words;

However, as it is well known in the art, Pedersen et al. teaches of tallying the frequency of words in respective paragraphs (C.2.lines 11-14-the frequency of each word is determined (summed) in order to be recorded, over the paragraphs). Therefore, at the time of the invention, it would have been obvious to combine Lawrence et al. with Pedersen et al. The motivation for doing so would have been to judge the relatedness of documents.

As per **claims 47, 52, and 57** Lawrence et al. and Pedersen et al. disclose all of the limitations of claim 46, upon which claim 47 depends. Lawrence et al. further discloses:

the step of determining includes stemming the common words of the associated paragraphs to a length that preserves their essential character while eliminating characters that convey little information about the word identity (C.4.lines 57-61-all words are stemmed including common words of the associated paragraphs).

Allowable Subject Matter

9. Claims 48-50, 53-55, and 58-60 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.
10. The following is a statement of reasons for the indication of allowable subject matter:

Regarding **claims 48, 53, and 58**, the closest prior art of Lawrence et al. reads on calculating respective distances of the sentences in the citing documents from respective citing instances of the cited document (C.2.lines 62-64-the “finding out where the article is cited” is interpreted as the calculated distance between the sentences and citing instance).

Prior art does not teach nor fairly suggest:

calculating respective initial content scores (ICS) for the sentences in the citing documents, based on the content words in the sentences.

calculating respective content scores (CS) for the sentences in the citing documents, based on at least the ICS and the distances.

11. **Claims 49, 50, 54, 55, 59, and 60** are allowable as they further limit their parent claims.

12. As allowable subject matter has been indicated, applicant's reply must either comply with all formal requirements or specifically traverse each requirement not complied with. See 37 CFR 1.111(b) and MPEP § 707.07(a).

Conclusion

13. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

14. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Lamont M Spooner whose telephone number is 703/305-8661. The examiner can normally be reached on 8:00 AM - 5:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Smits can be reached on 703/306-30113011. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Art Unit: 2654

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Technology Center 2600 receptionist whose telephone number is (703) 305-4700.

Ims
04/22/04



RICHMOND DORVIL
SUPERVISORY PATENT EXAMINER